UNITED STATES PATENT AND TRADEMARK OFFICE

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 10/723,049 | 11/26/2003 | Xian-Sheng Hua | MS1-1744US | 7823 |

22801          7590          10/06/2009
LEE & HAYES, PLLC
601 W. RIVERSIDE AVENUE
SUITE 1400
SPOKANE, WA 99201

| EXAMINER |
|---|
| GISHNOCK, NIKOLAI A |

| ART UNIT | PAPER NUMBER |
|---|---|
| 3715 | |

| NOTIFICATION DATE | DELIVERY MODE |
|---|---|
| 10/06/2009 | ELECTRONIC |

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

lhptoms@leehayes.com

| | **Application No.** | **Applicant(s)** |
|---|---|---|
| **Office Action Summary** | 10/723,049 | HUA ET AL. |
| | **Examiner** | **Art Unit** | |
| | NIKOLAI A. GISHNOCK | 3715 | |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on <u>16 July 2009</u>.
2a)☐ This action is **FINAL**.          2b)☒ This action is non-final.
3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) <u>1,5-8,10-15,17-19 and 21-45</u> is/are pending in the application.
    4a) Of the above claim(s) _____ is/are withdrawn from consideration.
5)☐ Claim(s) _____ is/are allowed.
6)☒ Claim(s) <u>1,5-8,10-15,17-19 and 21-45</u> is/are rejected.
7)☐ Claim(s) _____ is/are objected to.
8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.
10)☒ The drawing(s) filed on <u>26 November 2003</u> is/are: a)☒ accepted or b)☐ objected to by the Examiner.
    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
    a)☐ All   b)☐ Some *  c)☐ None of:
      1.☐ Certified copies of the priority documents have been received.
      2.☐ Certified copies of the priority documents have been received in Application No. _____.
      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
    * See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1)☒ Notice of References Cited (PTO-892)
2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3)☐ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.
5)☐ Notice of Informal Patent Application
6)☐ Other: _____.

## DETAILED ACTION

In response to Applicant's remarks filed 7/16/2009, claims 2-4, 7, 9, 16, & 20 are cancelled.

Claims 1, 5, 6, 8, 10-15, 17-19, & 21-52 are pending.

### Claim Rejections - 35 USC § 112

1.      The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

2.      The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

3.      Claim 15 is rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. It is unclear where support is found in the disclosure as originally filed for the limitation "in an event that both video and photograph are used, each photograph is regarded as a video shot". This is a New Matter rejection.

4.      Claim 1 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. With regards to the terms "if any," it is unclear in the claim how an attention/importance index of each frame is calculated if none of the recited factors are associated with the frame.

5.      Claim 38 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The term "and/or" is indefinite because it is not clear whether the language is inclusive or exclusive. Examiner suggests replacing "and/or" with "or".

6.      Claim 1 recites the limitation "the sub-shorts" in lines 19, 22, 23, 28, 30, 40, 42, 46, 48, 49, & 52.  There is insufficient antecedent basis for this limitation in the claim. Examiner suggests using either "sub-shots" or "sub-shorts" throughout the claims. Claim 14 recites wherein camera angles change, zoom, and pan the photograph. There is insufficient antecedent basis for this limitation in the claims. Claim 49 recites "the lower two layers" in lines 3-4. There is insufficient antecedent basis for this limitation in the claims.

7.      Claims 41-44 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention, because it is unclear whether the claim falls within the scope of 35 U.S.C. 112, sixth paragraph. A claim limitation will be presumed to invoke 35 U.S.C. 112, sixth paragraph, if it meets the following 3- prong analysis:

> **(A) the claim limitations must use the phrase "means for" or "step for; "**
> **(B) the "means for" or "step for" must be modified by functional language; and**
> **(C) the phrase "means for" or "step for" must not be modified by sufficient structure, material, or acts for achieving the specified function.**

See MPEP 2181. With respect to the third prong of this analysis, when a claim element uses language that generally falls under the step-plus-function format; however, 35 U.S.C. 112, sixth paragraph still does not apply when the claim limitation itself recites sufficient acts for performing the specified function; see *Seal-Flex*, 172 F.3d at 849, 50 USPQ2d at 1234. In the instant claims, the recited structural element of video analyzer performs the claimed means for defining and selecting visual content sub-shots, and the recited structural element of a lyric formatter performs the claimed means for timing delivery of syllables of lyrics. During

examination, applicants have the opportunity and the obligation to define their inventions

precisely, including whether a claim limitation invokes 35 U.S.C. 112, sixth paragraph. Thus, if

the phrase "means for" or "step for" is modified by sufficient structure, material or acts for

achieving the specified function, the USPTO will not apply 35 U.S.C. 112, sixth paragraph, until

such modifying language is deleted from the claim limitation. If a claim limitation does include

the phrase "means for" or "step for," that is, the first prong of the 3-prong analysis is met, but the

examiner determines that either the second prong or the third prong of the 3-prong analysis is

not met, then in these situations, the examiner must include a statement in the Office action

explaining the reasons why a claim limitation which uses the phrase "means for" or "step for" is

not being treated under 35 U.S.C. 112, sixth paragraph.


### Claim Objections

8.      Claim 14 is objected to because of the following informalities:  the word "zoom" is

misspelled.  Appropriate correction is required.

9.      Claim 28 is objected to under 37 CFR 1.75(c), as being of improper dependent form for

failing to further limit the subject matter of a previous claim.  Applicant is required to cancel or

amend the claim to place it in proper dependent form, or rewrite it in independent form.


### Claim Rejections - 35 USC § 103

10.     The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set
> forth in section 102 of this title, if the differences between the subject matter sought to be patented and
> the prior art are such that the subject matter as a whole would have been obvious at the time the
> invention was made to a person having ordinary skill in the art to which said subject matter pertains.
> Patentability shall not be negatived by the manner in which the invention was made.

11.     The factual inquiries set forth in *Graham* **v.** *John Deere Co.*, 383 U.S. 1, 148 USPQ 459

(1966), that are applied for establishing a background for determining obviousness under 35

U.S.C. 103(a) are summarized as follows:

  1.     Determining the scope and contents of the prior art.
  2.     Ascertaining the differences between the prior art and the claims at issue.
  3.     Resolving the level of ordinary skill in the pertinent art.
  4.     Considering objective evidence present in the application indicating obviousness
         or nonobviousness.

12.     Claims 1, 5, 6, 8, 10, 17, 18, 46, & 50 are rejected under 35 U.S.C. 103(a) as being

unpatentable over Stelovsky (US 5,782,692), hereinafter known as Stelovsky, in view of Golin

(US 5,990,980 A), hereinafter known as Golin, Osberger (US 6,670,963), hereinafter known as

Osberger, and Hansen et al. (US 2002/0038456 A1), hereinafter known as Hansen.

13.     Stelovsky teaches a processor-readable medium comprising processor- executable

instructions for personalizing karaoke, the processor-executable instructions comprising

instructions for performing a method (Column 1, Lines 54-67), the method comprising: obtaining

music (the multimedia game is retrieved and initialized, 4:32-40); obtaining lyrics corresponding

to the music from a file (textual track can be generated remotely and transmitted using

communications means, Column 14, Lines 20-24); selecting a visual content according to the

content, a user's preference, and a type of music with which the visual content is to be aligned

(In the "explore" mode, {the user} can click within the video window to start playing the music

video and click again to pause. The next click will resume the music video. This way all the

segments of the music video can be viewed in their natural sequence. The user can also click

on a segment in the segment bar to play the video starting with the selected segment. This

simple user interface allows the user to replay a segment, go to the next or previous segment or

choose an arbitrary segment, 8:46-59); segmenting music to produce a plurality of music (the

basic track consists of video display images and is synchronized with at least one other track

that consists of audio or text display, 3:31-35; The multimedia presentation is segmented with

respect to specific beginning and ending points of segments on the time axis, i.e. there are one

or more points of time that partition the time axis into time segments, 3:52-55); selecting sub-

shots from the plurality of sub-shots and aligning sub-shots with music sub-clips (multimedia

presentation track consisting of video, audio, and text display is segmented with respect to

specific beginning and ending points, Column 3, Lines 27-65), the aligning comprising:

automatically shortening one or more of the plurality of sub-shots to a length of a corresponding

music sub-clip from within the plurality of music sub-clips (the music video is synchronized with

a song's audio as well as the song's lyrics, and partitioned into time segments that correspond

to the song's phrases, Column 8, Lines 34-45); and resolving differences in the number of sub-

shorts and the number of music sub-clips (sets of choices {of the multimedia presentation} time

segments available to the user are specified and linked to the each of the time segments, 6:58-

66; the resolving of the differences in number is understood to merely mean not using some of

the time segments); coordinating delivery of the lyrics with the music using timing information

contained within the file (sections of the text track are linked to the time segments, Column 6,

Line 55; also Column 14, Lines 52-59 and Column 9, Lines 15-21); and displaying at least some

of the plurality of sub-shots as a background to lyrics associated with the plurality of music sub-

clips (the text can be superimposed on the video, Column 10, Lines 5-6) [Claim 1].

14.     What Stelovsky fails to teach is segmenting a visual content to produce a plurality of

sub-shots at a maximum peak of a frame difference curve, wherein the visual content presents

a story line, and repeating the dividing to result in sub-shots shorter than a maximum sub-shot

length, and producing a number of effects at transitions of the plurality of sub-shots [Claim 1].

However, Golin teaches the use of a Frame Dissimilarity Measure (FDM), which is the ratio of a

net dissimilarity measure and a cumulative dissimilarity measure of two consecutive frames

(Column 3, Line 65 to Column 4, Line 12). The processing of sub-shots uses the FDM to identify

transitions between shots in a video sequence, which appear as peaks in the FDM data

(Column 5, Lines 21-42). Golin further teaches where the FDM data is used to detect subshots

having transitions in them (5:54-62). The data analysis for the sub-shot dividing is a loop, which

starts with frames at the beginning of the video sequence and scans through the data to the

frames at the end of the sequence (Column 5, Lines 54-62). The length of the entire video

sequence is a maximum sub-shot length. Therefore, it would have been obvious to one of

ordinary skill in the art, at the time the invention was made, to have used the FDM peak analysis

of dividing sub-shots, as described in Golin, for the video segmenting used in Stelovsky, in order

to more effectively detect gradual transitions between sub-shots [Claim 1].

15.      What Stelovsky and Golin fail to explicitly teach is where the segmenting is repeated

until lengths of all sub-shots are shorter than a maximum of sub-shot length, the maximum of

sub-short length being a little longer in duration than the maximum of music sub-clips to

facilitate the sub-short being truncated to equal a length of an aligned music sub-clip in a next

step [Claim 1]. However, Stelovsky specifically teaches that the segmenting facilitates the

multimedia segments being truncated to equal the other segments in length (at 3:52-55);

Applicant has not disclosed that having the sub-shots be a "little longer" in duration than the

music sub-clips solves any stated problem or is for any particular purpose, when he intends to

shorten or "re-truncate" them again in a next step. Moreover, it appears that the length of the

sub-clips of Stelovsky or the Applicant's instant invention would perform equally well for

synchronizing the sub-clips with a video. Accordingly, it would have been obvious to one of

ordinary skill in the art, at the time the invention was made, to have modified Stelovsky such that

lengths of all sub-shots are shorter than a maximum of sub-shot length, the maximum of sub-

shot length being a little longer in duration than the maximum of music sub-clips, in light of

Golin, because such a modification would have been considered a mere design consideration, which fails to patentably distinguish over Stelovsky and Golin [Claim 1].

16.     What Stelovsky and Golin fail to teach is wherein the filtering of a plurality of sub-shots is according to importance or quality, the filtering sub-short from within the plurality of sub- shorts according to importance comprising: calculating an attention/importance index of each frame of the sub-shot based on a plurality of factors including object motion, camera motion, specific objects, and audio, if any, associated with the frame; calculating an attention/importance index of the sub-short by averaging the attention/importance index of each frame of the sub-short; and selecting the sub-shots by comparing the attention index of each sub-shot [Claim 1]. However, Osberger teaches giving areas of medium motion high importance (Column 7, Lines 10-21). Osberger also teaches that areas of low texture (quality) such as faces are strong attractors of attention (Column 8, Lines 40-54). The sub-shots that are high in "regions of interest", or attention attracting, are identified (filtered) as taught by Osberger (Column 2, Lines 24-41). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have used the methods of Osberger for filtering sub-shots based on attention indices such as importance to the camera and texture quality, in the karaoke video segmenting device of Stelovsky, in light of the teachings of Golin, in order to increase the entertainment value of the karaoke experience to a user [Claim 1].

17.     What Stelovsky, Golin, and Osberger fail to teach is and selecting sub-shots such that they are uniformly distributed along a time line of the visual content to preserve the story line of the visual content, or that the displaying comprises: merging the selected sub-shots into scenes by a plurality of grouping methods, the methods including: merging the sub-shorts by similarity and on a time-code or timestamp of the sub- shots [Claims 1, 23, 25 & 40]. However, Hansen teaches a system and method for automatically producing media content by creating video

subclips called "microchannels" by a "microchannel creator" that determines the desired

channel content based upon uniform distribution of video, video and audio, still images and

mosaics of different locations (The channel creator then accesses the individual clips from the

database and creates the continuous stream or "microchannel." The continuous stream is

defined by a concatenated stream of output, whether it be a series of images, video and audio,

or other forms of media; The microchannel creator makes the following decisions when creating

a microchannel: (i) what type of media should be sent at a given time (video, audio, image); (ii)

what triggers should be given priority, assuming multiple triggers are defined for the

microchannel; (iii) when advertising should be inserted into the video stream, and what

advertising should be provided; and (iv) when the database should be accessed for pre-

recorded clips that are not currently posted to the microchannel as new clips. The channel

creator runs via decision algorithms that are determined by the desired channel content for the

microchannel. This is best illustrated by example. Considering a hypothetical travel-related site,

the following type of microchannel might be desired: (i) commercials should be presented once

per minute in ten second maximum durations; (ii) uniform distribution of video, video and audio,

still images and mosaics of different locations; (iii) emphasis on video content using activity

triggers on beach cams and urban cams; (iv) emphasis on mosaic content using periodic

triggering without motion for panoramic cameras; (v) emphasis on still image content for interior

cameras, such as restaurant cameras; (vi) live, real-time clips during daylight hours; and (vii)

pre-recorded clips during night hours when beach activity has ceased, Para. 0085-88). As best

understood, Hansen teaches selecting "microchannels" uniformly from a source. The

"microchannels" are understood to be a grouping of subshots, having a time-code or time-stamp

in close temporal proximity, and with image similarity based on location. The "microchannel

creator" of Hanson would be used in the device of Stelovsky to uniformly select video and

photographic content. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to selecting sub-shots such that they are uniformly distributed within the video and grouping, as taught by Hansen, in the device of Stelovsky, in light of Golin and Osberger, in order to automatically produce and distribute media content to a targeted audience, for providing more interesting and representative content [Claim 1].

18.     What Stelovsky, Golin, and Hansen fail to teach is wherein filtering the plurality of sub-shots comprises instructions for: examining color entropy within each of the plurality of sub-shots to detect motion more than a threshold indicating interest and less than a threshold indicating low camera and/or object movement; and selecting sub-shots having acceptable motion and/or color entropy scores [Claim 5]. However, Osberger teaches segmenting frames into regions based upon both color and luminance (Column 2, Lines 24-41). The term entropy is taken to mean Information Entropy or Shannon Entropy, which refers to a measure of uncertainty associated with a random variable. Thus, referring to lossless data compression, the color entropy would refer to an average minimum number of bits needed to communicate a color value. Osberger teaches using an algorithm to segment an image into homogeneous regions using color information, to generate the spatial importance map (Column 3, Lines 6-15). Osberger also teaches that, if the spatial importance map is too noisy from frame to frame, a temporal smoothing operation is performed, and a temporal importance map is generated (Column 6, Line 66 to Column 7, Line 37). The temporal importance map is calculated using adaptable thresholds because the amount of motion varies greatly across different scenes. Osberger also teaches identifying sub-shots with regions of interest by using the spatial and temporal interest maps in order to produce an adaptive segmentation model (Column 8, Lines 58-67), for segmenting video scenes. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have incorporated the color entropy

detection, then the camera motion detection of Osberger with the segmentation of karaoke video as described by Stelovsky, in light of Golin and Hansen, in order to attract the interest of a karaoke user more effectively [Claim 5].

19.      What Stelovsky, Wang, Hansen, and Umeda further fail to teach is wherein filtering the plurality of sub-shots according to importance comprises instructions for evaluating frames within a sub-shot according to attention indices, and averaging the attention indices for the frames to determine if the sub-shot should be included or excluded [Claim 6]. However, Osberger teaches identifying and adaptively segmenting frames of video based upon an attention model, AKA total importance map, composed by linear weighting of the spatial and temporal importance maps (Column 2, Lines 24-41). It is inherent that averaging is merely linear weighting with a weight factor of one. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have utilized the averaging of the attention indices of Osberger to select frames of importance, for use in the karaoke system of Stelovsky, in light of Golin and Hansen, in order to adapt the attention model for a variety of different types of video sub-shots, while accurately determining regions of interest in the videos [Claim 6].

20.      What Stelovsky , Golin, and Osberger further fail to teach is wherein each sub-shot comprises a segment of video of at least a predetermined length based on the length of the music sub-clips and segmented based on a magnitude of difference between adjacent frames [Claim 8]. However, Hansen teaches a system and method for automatically producing media content, in which the clip has a predetermined minimum length {one still frame}, based on detected trigger events in the clip (A "clip" may be defined as a duration of time when the triggers that are set for the capture system are activated--such as when there is motion in the scene and the trigger is set to a basic motion cue. The clip preferably ends when the trigger event is no longer detected or when a certain time period expires, although other more

sophisticated methods for trigger intervals may also be utilized. Once a clip is delineated, the

content is generated. At a minimum, the content includes one still image that represents the

trigger event in action. For example, 15 seconds out of one minute of captured content may be

identified as qualifying content, Para. 0043). The minimum length of a segmented sub-shot of

Stelovsky would merely be pre-determined, either one frame long, or based on a trigger such as

the detected events in the clip, as taught by Hansen. Therefore, it would have been obvious to

one of ordinary skill in the art, at the time the invention was made, to have each sub-shot of

Stelovsky comprise a segment of video of at least a predetermined length based on the length

of the music sub-clips and segmented based on a magnitude of difference between adjacent

frames, as taught by Hansen, in light of the teachings of Golin and Osberger, in order to

automatically produce media content pleasing to a user's senses [Claim 8].

21.      What Stelovsky, Golin, and Hansen further fail to teach is wherein a visual content

analyzer is configured to select from the sub-shots according to ranked importance, gauged by

detection of color entropy, object motion, camera motion, or of a face within the sub-shot

[Claims 10 & 32]. However, Osberger teaches selecting or filtering sub-shots by color

information (Column 3, Lines 6-15), by camera or object motion (Column 7, Lines 7-37), or by

specific objects, including faces, in a sub-shot (Column 8, Lines 40-54). Therefore, it would have

been obvious to one of ordinary skill in the art, at the time the invention was made, to have used

the various color, motion, and object detection in the video sub-shots, as described by

Osberger, in the personalized karaoke system on Stelovsky, in light of Golin and Hansen, in

order to improve the prediction of visual importance of a sub-shot [Claim 10].

22.      What Stelovsky, Golin, Osberger, and Hansen further fail to explicitly teach is wherein

the segmenting music comprises instructions for bounding the sub-clip's length according to:

minimum length = min(max(2*tempo,2),4) and maximum length = minimum length+2 [Claim 17],

or establishing the music sub-clip's length within a range of 3 to 5 seconds [Claim 18]. However,

Applicant has not disclosed that having (min(max(2\*tempo,2),4) < length <

min(max(2\*tempo,2),4)+2) or (3 < length < 5) seconds solves any stated problem or is for any

particular purpose. Moreover, it appears that the arbitrary length of the sub-clips of Stelovsky or

the Applicant's instant invention would perform equally well for synchronizing the sub-clips with

a video. Accordingly, it would have been obvious to one of ordinary skill in the art, at the time

the invention was made, to have modified Stelovsky such that the music sub-clips had a rigid

minimum and maximum length, in light of Golin, Osberger, and Hansen, because such a

modification would have been considered a mere design consideration, which fails to patentably

distinguish over Stelovsky, Golin, Osberger, and Hansen [Claims 17 & 18].

23.     Stelovsky teaches wherein the number of effects at transitions of the plurality of sub-

shots are selected randomly in a plurality of specific effect sets or determined by a style (SAS

also supports the identification of additional resources associated with each segment and each

event. Examples of such resources include additional discrete or continuous media tracks, such

as icons, still images, audio, motion video tracks and hypertext links leading to information

associated with the segment or the event. These additional resources can be independent,

constitute a predefined sequence, or be tied to a time point in between the start and end point of

the segment {respective the segment to which the event belongs}, 7:51-62; it is understood that

the effect transitions are thus part of a style respective to the segment to which the event

belongs) [Claim 50].


24.     Claim 11 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky,

Golin, Osberger, and Hansen, as applied to claim 1 above, and further in view of Haitsma et al.

(US 2002/0178410 A1), hereinafter known as Haitsma.

25.    Stelovsky, Golin, Osberger, and Hansen teach all the features of claim 1 as

demonstrated above. What Stelovsky, Golin, Osberger, and Hansen fail to teach is wherein the

selecting uniformly distributed sub-shots comprises evaluating a normalized entropy of the sub-

shots along a time line of video from which the sub-shots are obtained [Claim 11]. However,

Haitsma teaches a hashing method for indexing video clips in a database, in which a normal

distribution is calculated for video clips to determine whether they are different quality versions

of the same content (Two 3 seconds audio clips {or two 30-frame video sequences} are

declared similar if the Hamming distance between the two derived hash blocks $H.sub.1$ and

$H.sub.2$ is below a certain threshold T. This threshold T directly determines the false positive

rate $P.sub.f$, i.e. the rate at which two audio clips/video sequences are incorrectly declared

equal {i.e. incorrectly in the eyes of a human beholder}: the smaller T, the smaller the probability

$P.sub.f$ will be. On the other hand, a small value T will negatively effect the false negative

probability $P.sub.n$, i.e. the probability that two signals are `equal`, but not identified as such. In

order to analyze the choice of this threshold T, we assume that the hash extraction process

yields random i.i.d. {independent and identically distributed} bits. The number of bit errors will

then have a binomial distribution with parameters {n,p}, where n equals the number of bits

extracted and p(=0.5) is the probability that a `0` or `1` bit is extracted. Since n

{32.times.256=8192 for audio, 32.times.30=960 for video} is large in our application, the

binomial distribution can be approximated by a normal distribution with a mean $.mu.=np$ and

standard deviation $.sigma.=$[square root][square root over (np(1-p))], Para. 0041). This is

understood to be a normalized entropy in the sense that the normal video quality is used to

determine the similarity of sub-shots. Such a method would be used in the system and method

of Stelovsky to determine whether a video clip or photograph duplicates the content of another

except in quality. Therefore, it would have been obvious to one of ordinary skill in the art, at the

time the invention was made, to have selected a uniform distribution of sub-shots along a

timeline, as taught by Hansen, by analyzing the normalized entropy of the sub-shots, as taught

by Haitsma, in light of the teachings of Golin, Osberger, and Hansen, in order to avoid the non-

uniform selection of duplicate sub-shot content in sub-shots that have distinct data

representations due to differing quality [Claim 11].

26.     Claims 12-15, & 47-49 are rejected under 35 U.S.C. 103(a) as being unpatentable over

Stelovsky, Wang, Golin, Osberger, and Hansen, and further in view of Geigel et al. (US

2002/0122067 A1), hereinafter known as Geigel.

27.     Stelovsky, Golin, Osberger, and Hansen teach all the features as demonstrated in the

rejection of claim 1 above. Osberger teaches a visual analyzer (image processing algorithm) to

detect an attention area within a photograph (Column 2, Lines 24-41), and wherein motion

vectors are used by camera motion estimation algorithm to determine pan and zoom in a frame

(Column 7, Lines 22-37). What Stelovsky, Golin, Osberger, and Hansen fail to explicitly teach is

wherein the instructions for segmenting visual content includes assigning photographs to be

sub-shots [Claim 12], instructions for assigning photographs includes converting at least one

photograph to video, where the camera angles change, zoom, and pan the photograph [Claim

14], wherein the visual content comprises one or more home videos or photographs in digital

formats, where photographs are regarded as video shots [Claim 15]. However, Geigel teaches a

layout generator for digital images (Para. 0010), including photographs or video clips (Para.

0055), which converts the images into a video (output is Picture CD media or other photo

delivery media, Para. 0057). It is inherent that a series of images displayed during a progression

of time is a video. Geigel also teaches, in photography terms rather than videography terms,

panning the images (auto-cropping, Para. 0057) and zooming the images (scaling, Para. 0122).

Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have assembled and converted photos to video, as taught by Geigel, for the background video in the entertainment system of Stelovsky, in light of Golin, Osberger, and Hansen, in order to automate the layout of the background in a manner pleasing to the user; and to create a photo to video sub-shot based on a detected attention area, including panning and zooming, in light of the teachings of Osberger and Geigel, in the entertainment system of Stelovsky, in light of Wang and Hansen, in order to further refine the content information of an image by focusing on the attention-attracting elements in the photo to video, when used as the background for karaoke entertainment [Claims 12, 14, & 15].

28.     What Stelovsky, Golin, Osberger, and Hansen further fail to teach is wherein a visual content analyzer is configured with instructions for assigning photographs includes instructions for: rejecting photographs having problems with quality, and rejecting a similar group of photographs when one within the group has been selected [Claim 13]. However, Geigel teaches performing detection of dud images and duplicate images prior to being submitted to the layout system (Para. 0061). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have not selected dud or duplicate images when creating the background image layout, as shown by Geigel, when implementing the entertainment system of Stelovsky, in light of Golin, Osberger, and Hansen, in order to necessitate the minimal input from the user when assembling images aesthetically pleasing to the user [Claim 13].

29.     What Stelovsky further fails to teach is wherein the one or more photographs are grouped into three tiers including: a date that the photograph is taken, a scene within the photograph, and whether the photo is a member of a group of very similar photographs, wherein the scene represents a group of photographs that, while not as similar as those which fall under

the group of very similar photos, are taken at a same time and place [Claim 47], wherein the

date and scene are used to determine the number of effects at transition of the one or more

photos and photos fall within a group of very similar photos are filtered out [Claim 48], or

wherein the photographs are firstly grouped into a top-tier based on the date, then a hierarchical

clustering algorithm with different similarity thresholds is used to group the lower two layers,

wherein the photographs with a lower degree of similarity are grouped together as the scene

[Claim 49]. However, Geigel teaches grouping photographs by chronology, (Para. 0063), event

and sub-event (Para. 0059-60), emphasis (Para. 0083), and unity (Para. 0084). The transitions

in the karaoke system of Stelovksy would merely be selected based on the type of grouping of

each set of photographs. Therefore, it would have been obvious to one of ordinary skill in the

art, at the time the invention was made, to have used the photographic tiers of grouping of

Geigel in order to determine transition effects of Stelovsky, in order that the transitions between

similar photographs is the same, but different from scene transitions, to generate a more

pleasing karaoke for a viewer [Claims 47-49].


30.     Claims 19 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky,

Golin, Osberger, and Hansen, and further in view of Bloom et al. (US 2005/0042591 A1),

hereinafter known as Bloom.

31.     Stelovsky, Hansen, Golin, and Osberger teach all the features as demonstrated above in

the rejections of claims 1, 18, 25, & 40 above, including wherein the lyric formatter is configured

to consume a file detailing timing of the lyrics (the textual track can be generated remotely and

transmitted by communication means, digitally, using a software program, Column 14, Lines 14-

24; the digital textual track used for the karaoke is inherently a file to be "consumed" or used).

Stelovsky teaches wherein evaluation of output can involve differences in pronunciation patterns

and any processes involved in generating speech (Column 14, Lines 52-59). What Stelovsky, Hansen, Golin, and Osberger fail to teach is wherein segmenting the music comprises a lyric formatter configured with instructions for establishing boundaries for the music sub-clips at sentence breaks in lyrics [Claim 19]. However, Bloom teaches automatically synchronizing sound to images, wherein lyric segmentation may be syllable by syllable (line can be a single word or sound) or a sentence (Para. 0139). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have segmented the music of the karaoke system of Stelovsky, in light of the syllable and sentence boundaries of the lyrics as taught by Bloom, in light of Hansen, Golin, and Osberger, in order to synchronize the song with a user's lip movements on the accompanying video display [Claim 19].

32.    Claim 21 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in view of Golin, Osberger, and Hansen, and further in view of Tsai (US 6,572,381 B1), hereinafter known as Tsai.

33.    Stelovsky, Golin, Osberger, and Hansen, teach all the features as demonstrated above in the rejections of claim 1 above. What Stelovsky, Golin, Osberger, and Hansen fail to teach is wherein obtaining the lyrics comprises instructions for sending the file over a network to a karaoke device as part of a pay-for-play service [Claim 21]. However, Tsai teaches a plurality of karaoke terminals connected to a host computer via a network (communications line) that delivers lyric data (Column 8, Lines 48-61). Tsai teaches a karaoke system shares the source data as part of a pay service (Column 2, Lines 48-56; also Column 20, Line 52 to Column 21, Line 56). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have sent the lyrics file over a network in conjunction with a pay-for-play service, as taught by Tsai, in the karaoke system of Stelovsky, in light of Golin, Osberger, and

Hansen, in order to offer commercial messages with updated custom content to a subscriber of a karaoke service [Claim 21].

34.    Claim 22 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in view of Golin, Osberger, and Hansen, and further in view of Tashiro et al. (US 5,703,308), hereinafter known as Tashiro.

35.    Stelovsky, Golin, Osberger, and Hansen teach all the features as demonstrated above in the rejections of claim 1 above. What Stelovsky, Golin, Osberger, and Hansen fail to teach is wherein the processor-readable medium comprises instructions for: querying a database of songs by humming a portion of a desired song; and selecting the desired song from among a number of possibilities suggested by an interface to the database [Claim 22]. However, Tashiro teaches a karaoke device having database of songs (music data storage device with a plurality of entry songs stored in a data table, Column 1, Line 54 to Column 2, Line 3), wherein the database is queried by humming a song (key melody patterns which represent a desired song are input by voice, Column 3, Lines 10-14) and selecting the desired song through an interface (music selection is made from top 10 matching entries, Column 7, Lines 48-67). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, in the karaoke system of Stelovsky, to search and select a desired song from a database by humming, as taught by Tashiro, in light of Golin, Osberger, and Hansen, in order to select a song even if neither the artist nor the title of the song is known [Claim 22].

36.    Claims 23-25, 28, 29, 32, 33, 40, 41, 43, & 46 are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, Golin, Osberger, and Hansen, and further in view of Wang, (US 2002/0133764 A1), hereinafter known as Wang.

37.    Stelovsky teaches a processor-readable medium comprising instructions for providing

lyrics for integrating lyrics, music, and video content suitable for karaoke, comprising

instructions for: receiving a request for a file associated with a specific song (clicking on a word

in the text track, Column 14, Lines 42-48), wherein the file comprises music, lyrics, and timing

values (The time-dependent sequence is composed of tracks that are synchronized with respect

to a common time axis {hereinafter "multimedia presentation"}. The basic track consists of video

display images and is synchronized with at least one other track that consists of audio or text

display, 3:31-35; The multimedia presentation is segmented with respect to specific beginning

and ending points of segments on the time axis, i.e. there are one or more points of time that

partition the time axis into time segments, 3:52-55), fulfilling the request by sending the file

associated with the specified song (connection is established with a remote on-line service,

search query initiated, and results are displayed, Column 14, Lines 42-48), segmenting visual

content to produce a plurality of sub-shots of a length corresponding to the music sub-clips

(multimedia presentation track consisting of video, audio, and text display is segmented with

respect to specific beginning and ending points, Column 3, Lines 27-65), and outputting the

plurality of music sub-clips together with corresponding sub-shots of visual content, which is

configured as a background to the lyrics associated with the music sub-clips ("Karaoke Game"

presentation has synchronized video and instrumental sound tracks, Column 9, Lines 15-21; the

text can be superimposed on the video, Column 10, Lines 5-6) [Claim 23].

38.    Stelovsky teaches a personalized karaoke device, comprising: a music analyzer

configured to create music sub-clips of varying lengths according to a song (Segmentation

Authoring System {SAS} facilitates the identification of points in time where a segment starts

and ends, Column 5, Line 62 to Column 6, Line 2; multimedia presentation track consisting of

video, audio, and text display is segmented with respect to specific beginning and ending points,

Column 3, Lines 27-65); a visual content analyzer configured to define and select visual content sub-shots (Using SAS, the author partitions the multimedia presentation into time segments according to predominant time units, e.g., measures of song, sound bites, or action sequences in a movie, Column 6, Lines 51-54); a lyric formatter configured to time delivery of syllables of lyrics of the song (evaluation feedback of user's input includes visualization of differences in pronunciation patterns, processes involved in generating {human} speech, such as positions of the tongue and air flow patterns, Column 14, Lines 52-59; it is inherent that the speech analysis as disclosed could recognize syllables and sentences, which are pronunciation patterns; sections of the text track are linked to the time segments, Column 6, Line 55); and a composer configured to assemble the music sub-clips with the visual content sub-shots, and configured to adjust the length of the sub-shots to correspond to the music sub-clips, and to superimpose the syllables of the lyrics of the song over the sub-shots ({SAS} sections of a text track and additional media resources are linked to the time segments, Column 6, Lines 55-57) [Claim 25].

39.      Stelovsky teaches an apparatus, comprising: means for creating music sub-clips according to a song, and means for defining and selecting visual content sub-shots (multimedia presentation track consisting of video, audio, and text display is segmented with respect to specific beginning and ending points, Column 3, Lines 27-65); means for timing delivery of syllables of lyrics of the song (sections of the text track are linked to the time segments, Column 6, Line 55; the text can be superimposed on the video, Column 10, Lines 5-6, also Column 14, Lines 52-59 and Column 9, Lines 15-21); means for automatically assembling the music sub-clips with the visual content sub-shots and adjusting the length of the sub-shots to correspond to the length of the music sub-clips (the music video is synchronized with a song's audio as well as the song's lyrics, and partitioned into time segments that correspond to the song's phrases, Column 8, Lines 34-45), to superimpose the syllables of the lyrics of the song over the sub-

shots (While the song is playing, the corresponding phrases are highlighted in the lyrics field. If

necessary, the lyric's field is automatically scrolled to reveal the current phrase, Column 8, Lines

34-45) [Claim 40].

40.      What Stelovsky fails to teach is where the segmenting of music to produce a plurality of

music sub-clips establishes boundaries between the music sub-clips at beat positions within the

music, the beat positions being located according to a rhythm or a tempo of the music [Claims

23, 25, & 40], and wherein each sub-clip has a duration that is a function of song tempo [Claim

28], or at onset positions within the music when beat positions are not obvious during a portion

of the music, the onset positions being initiations of distinguishable tones of the portion of the

music [Claim 46]. However, Wang teaches a method of detecting beats in a music stream (Beat

is defined in the relevant art as a series of perceived pulses dividing a musical signal into

intervals of approximately the same duration. Beat detection can be accomplished by any of

three methods. The preferred method uses the variance of the music signal, which variance is

derived from decoded Inverse Modified Discrete Cosine Transformation (IMDCT) coefficients.

The variance method detects primarily strong beats. The second method uses an Envelope

scheme to detect both strong beats and offbeats. The third method uses a window-switching

pattern to identify the beats present. The window-switching method detects both strong and

weaker beats. In one embodiment, a beat pattern is detected by the variance and the window

switching methods. The two results are compared to more conclusively identify the strong beats

and the offbeats, Para. 0070-0074; see also Figure 7, the numbered delta functions are

understood to be detected beats), and segmenting the music stream at beat boundaries (A

normal, error-free audio transmission is represented in the top graph {of Figure 6} by a first and

second beat-to-beat interval waveform. The first waveform includes a first beat and the audio

information up to a second beat. Similarly, the second waveform includes the second beat and

the audio information up to a third beat; In accordance with the method of the present invention, a replacement waveform, including a replacement beat, is copied from the first beat and the first waveform; and is substituted for the missing audio segment in the time interval T1 to T2, as shown in the bottom graph; all at Para. 0058-0069; see also Figure 6). The beat intervals are taught by Wang to be a function of song tempo (the beat-to-beat interval is replaced by the audio data frames from a corresponding beat-to-beat interval in a preceding 4/4 bar. Most popular music has a rhythm period in 4/4 time, Para. 0067; 4/4 time is understood to be a tempo). Any of the three methods taught by Wang would be used to detect beats in a music clip, and Wang's method of copying and pasting music waveforms segmented by at beat positions would be used to align video, still pictures, music, and lyrics along those boundaries, in the manner as taught by Stelovsky. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have used Wang's methods of segmenting of music to produce a plurality of music sub-clips, establishing boundaries between the music sub-clips at beat positions within the music, with the methods of Stelovsky for integrating lyrics, music, and video content suitable for karaoke, in order to exploit the beat pattern of music signals to improve the presentation of music when transferred over a network [Claims 23, 25, 28, 40, & 46].

41.     Stelovsky teaches wherein obtaining lyrics comprises instructions for sending the file over a network to a karaoke device (textual track can be generated remotely and transmitted using communications means, Column 14, Lines 20-24; on-line services provide downloading of files, e.g. Internet, Column 6, Lines 49-50) [Claim 24].

42.     Stelovsky teaches wherein the visual content analyzer is configured to segment video into sub-shots (Column 6, Lines 51-54) [Claim 29].

43.     What Stelovsky, Wang, Golin, and Hansen further fail to teach is wherein a visual

content analyzer is configured to select from the sub-shots according to ranked importance,

gauged by detection of color entropy, object motion, camera motion, or of a face within the sub-

shot [Claim 32]. However, Osberger teaches selecting or filtering sub-shots by color information

(Column 3, Lines 6-15), by camera or object motion (Column 7, Lines 7-37), or by specific

objects, including faces, in a sub-shot (Column 8, Lines 40-54). Therefore, it would have been

obvious to one of ordinary skill in the art, at the time the invention was made, to have used the

various color, motion, and object detection in the video sub-shots, as described by Osberger, in

the personalized karaoke system on Stelovsky, in light of Wang, Golin and Hansen, in order to

improve the prediction of visual importance of a sub-shot [Claim 32].

44.     What Stelovsky, Wang, Golin, and Hansen fail to teach is wherein the visual content

analyzer is configured to filter out sub-shots having low image quality, as measured by low

entropy and low motion intensity [Claim 33]. However, Osberger teaches segmenting frames

into regions based upon both color and luminance (Column 2, Lines 24-41). The term entropy is

taken to mean Information Entropy or Shannon Entropy, which refers to a measure of

uncertainty associated with a random variable. Thus, referring to lossless data compression, the

color entropy would refer to an average minimum number of bits needed to communicate a

color value. Osberger teaches using an algorithm to segment an image into homogeneous

regions using color information, to generate the spatial importance map (Column 3, Lines 6-15).

Osberger also teaches that, if the spatial importance map is too noisy from frame to frame, a

temporal smoothing operation is performed, and a temporal importance map is generated

(Column 6, Line 66 to Column 7, Line 37). The temporal importance map is calculated using

adaptable thresholds because the amount of motion varies greatly across different scenes.

Osberger also teaches identifying sub-shots with regions of interest by using the spatial and

temporal interest maps in order to produce an adaptive segmentation model (Column 8, Lines

58-67), for segmenting video scenes. Therefore, it would have been obvious to one of ordinary

skill in the art, at the time the invention was made, to have incorporated the color entropy

detection, then the camera motion detection of Osberger with the segmentation of karaoke

video as described by Stelovsky, in light of Wang, Golin, and Hansen, in order to attract the

interest of a karaoke user more effectively [Claim 33].

45.     Stelovsky teaches wherein the means for defining and selecting visual content sub-shots

is a video analyzer configured to segment video into sub-shots (Using SAS, the author partitions

the multimedia presentation into time segments according to predominant time units, e.g., action

sequences in a movie, Column 6, Lines 51-54) [Claim 41].


46.     Claim 26 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in

view of Wang Golin, Osberger, and Hansen, and further in view of Trovato et al. (US 7,058,889

B2), hereinafter known as Trovato.

47.     Stelovsky, Wang, Golin, Osberger, and Hansen teach all the features as demonstrated

above in the rejections of claim 25. What Stelovsky, Wang, Golin, Osberger, and Hansen fail to

teach wherein the music analyzer is configured to segment the song with a strong onset

between each of the music sub-clips [Claim 26]. However, Trovato teaches locating transition

points for a music segmentation scheme by onset break detection (Column 7, Lines 33-51; also

Figure 6). It is inherent from Figure 6 that weak onset breaks are not used as transition points.

Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention

was made, to have analyzed the music used in the karaoke system of Stelovsky with the onset

break detection method defined in Trovato, in light of Wang, Golin, Osberger, and Hansen, in

order to automatically synchronize the music with the background video consistent with human

perception [Claim 26].

48.    Claim 27 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in

view of Wang, Golin, Osberger, and Hansen, and further in view of Kondo (US 6,232,540 B1),

hereinafter known as Kondo.

49.    Stelovsky, Wang, Golin, Osberger, and Hansen teach all the features as demonstrated

above in the rejections of claim 25. What Stelovsky, Wang, Golin, Osberger, and Hansen fail to

teach is wherein a music analyzer is configured to segment the music automatically, comprising

instructions for: establishing boundaries for the music sub-clips with a beat position between

each of the music sub-clips [Claim 27]. However, Kondo teaches establishing boundaries

(positions) for music sub-clips (rhythm sound source signals) at beat positions within the music

(positions of attacks in the rhythm sounds, Abstract). Therefore, it would have been obvious to

one of ordinary skill in the art, at the time the invention was made, to have divided the music

sub-clips at beat positions within the music, as shown in Kondo, for use in the karaoke system

of Stelovsky, in light of Wang, Golin, Osberger, and Hansen, in order to avoid occurrences of

rhythm disorder in the rhythm sounds [Claim 27].

50.    Claims 30 & 42 are rejected under 35 U.S.C. 103(a) as being unpatentable over

Stelovsky, Wang, Golin, Osberger, and Hansen, in view of Borden, IV et al. (US 2003/0200105

A1), hereinafter known as Borden IV.

51.    Stelovsky, Wang, Golin, Osberger, and Hansen teach all the features of claims 25 & 40

above. What Stelovsky, Wang, Golin, Osberger, and Hansen fail to teach is where the video

analyzer or visual content analyzer is configured to access folders of home videos and

photographs containing content from which the sub-shots are derived [Claims 30 & 42].

However, Border IV teaches a video analyzer (user's data processing device) which can access

folders of a customer's video or photographs (MY PHOTOS homepage document, containing a

user's uploaded images or video, Para. 0016-0017). Therefore, it would have been obvious to

one of ordinary skill in the art, at the time the invention was made, to have accessed a user's

personal video and photo content for generating the sub-shots, in the karaoke device of

Stelovsky, in light of Wang, Golin, Osberger, and Hansen, in order to attract potential customers

to receive services by hosting their personal data [Claims 30 & 42].

52.     Claims 31, 34-38, & 43 are rejected under 35 U.S.C. 103(a) as being unpatentable over

Stelovsky, Wang, Golin, Osberger, and Hansen, and further in view of Geigel et al. (US

2002/0122067 A1), hereinafter known as Geigel.

53.     Stelovsky, Wang, Golin, Osberger, and Hansen teach all the features as demonstrated

in the rejection of claim 1 above. What Stelovsky, Wang, Golin, Osberger, and Hansen fail to

explicitly teach wherein a visual content analyzer is configured to assemble still photographs,

each of which is a sub-shot [Claim 31], and wherein the visual content analyzer is configured to

define sub-shots from visual content comprising photographic and video content [Claim 34].

However, Geigel teaches a layout generator for digital images (Para. 0010), including

photographs or video clips (Para. 0055), which converts the images into a video (output is

Picture CD media or other photo delivery media, Para. 0057). It is inherent that a series of

images displayed during a progression of time is a video. Therefore, it would have been obvious

to one of ordinary skill in the art, at the time the invention was made, to have assembled and

converted photos to video, as taught by Geigel, for the background video in the entertainment

system of Stelovsky, in light of Wang, Golin, Osberger, and Hansen, in order to automate the layout of the background in a manner pleasing to the user [Claims 31, & 34].

54.    What Stelovsky, Wang, Golin, Osberger, and Hansen fail to teach is wherein a visual content analyzer is configured to reject photographs of low quality by detecting over and under exposure, overly homogeneous images, and blurred images [Claim 35]. Osberger teaches a visual analyzer (image processing algorithm) to detect overexposure and underexposure (contrast), overly homogeneous images (homogeneous regions, Column 3, Lines 6-15), and blurred images (areas of very high motion, Column 7, Lines 10-26). What Stelovsky, Wang, Hansen, and Osberger fail to teach is wherein the visual content analyzer rejects photographs which are underexposed, overexposed, overly homogeneous, or blurred [Claim 35]. However, Geigel teaches selection of the best image (Para. 0057). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have rejected images which are underexposed, overexposed, overly homogeneous, or blurred, in light of the teachings of Osberger and Geigel, in the entertainment system of Stelovsky, in light of Wang, Golin, and Hansen, in order to discriminate images to present highly desirable visuals to a karaoke user [Claim 35].

55.    What Stelovsky, Wang, Golin, Osberger, and Hansen further fail to teach is wherein a visual content analyzer is configured to organize photographs by the date of exposure and scene, thereby obtaining photographs having a relationship [Claim 36]. However, Geigel teaches organizing the images (page layout algorithm, Para 0059) by date of exposure (chronology of the images, Para. 0063) and scene (event clustering, Para. 0060). It is inherent that all the photographs would thus be related by a date range or event group. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have organized the images to the extent provided by Geigel, is the operation of the

entertainment system of Stelovsky, in light of Wang, Golin, Osberger, and Hansen, in order to

distribute the photographs automatically according to an algorithm that valued a user-pleasing

arrangement [Claim 36].

56.     What Stelovsky, Wang, Golin, Osberger, and Hansen further fail to teach is rejecting a

similar group of photographs when one within the group has been selected [Claim 37].

However, Geigel teaches performing detection of dud images and duplicate images prior to

being submitted to the layout system (Para. 0061). Therefore, it would have been obvious to

one of ordinary skill in the art, at the time the invention was made, to have not selected dud or

duplicate images when creating the background image layout, as shown by Geigel, when

implementing the entertainment system of Stelovsky, in light of Wang, Golin, Osberger, and

Hansen, in order to necessitate the minimal input from the user when assembling images

aesthetically pleasing to the user [Claim 37].

57.     What Stelovsky, Wang, Golin, and Hansen further fail to teach is wherein the means for

defining and selecting visual content sub-shots is a video analyzer configured for: detecting an

attention area within a photograph; and creating a photo to video sub-shot based on the

attention area, wherein the video includes panning and zooming [Claims 38 & 43]. Osberger

teaches a visual analyzer (image processing algorithm) to detect an attention area within a

photograph (Column 2, Lines 24-41), and wherein motion vectors are used by camera motion

estimation algorithm to determine pan and zoom in a frame (Column 7, Lines 22-37). What

Stelovsky, Wang, Golin, Hansen, and Osberger fail to teach is wherein photo to video sub-shot

includes panning and zooming. However, Geigel teaches, in photography terms rather than

videography terms, panning the images (auto-cropping, Para. 0057) and zooming the images

(scaling, Para. 0122). Therefore, it would have been obvious to one of ordinary skill in the art, at

the time the invention was made, to created a photo to video sub-shot based on a detected

attention area, including panning and zooming, in light of the teachings of Osberger and Geigel, in the entertainment system of Stelovsky, in light of Wang, Golin, and Hansen, in order to further refine the content information of an image by focusing on the attention-attracting elements in the photo to video, when used as the background for karaoke entertainment [Claims 38 & 43].

58.      Claims 39 & 44 are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, Wang, Golin, Osberger, and Hansen, and further in view of Bloom et al. (US 2005/0042591 A1), hereinafter known as Bloom.

59.      Stelovsky, Wang, Hansen, Golin, and Osberger teach all the features as demonstrated above in the rejections of claims 25 & 40 above, including wherein the lyric formatter is configured to consume a file detailing timing of the lyrics (the textual track can be generated remotely and transmitted by communication means, digitally, using a software program, Column 14, Lines 14-24; the digital textual track used for the karaoke is inherently a file to be "consumed" or used). Stelovsky teaches wherein evaluation of output can involve differences in pronunciation patterns and any processes involved in generating speech (Column 14, Lines 52-59). What Stelovsky, Wang, Hansen, Golin, and Osberger fail to teach is consuming a file detailing timing of each syllable and each sentence of the lyrics [Claims 39 & 44], and for rendering the lyrics syllable by syllable [Claim 44]. However, Bloom teaches automatically synchronizing sound to images, wherein lyric segmentation may be syllable by syllable (line can be a single word or sound) or a sentence (Para. 0139). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have segmented the music of the karaoke system of Stelovsky, in light of the syllable and sentence boundaries of the lyrics as taught by Bloom, in light of Wang, Hansen, Golin, and Osberger, in order to

synchronize the song with a user's lip movements on the accompanying video display [Claims 39 & 44].


60.     Claim 45 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, Wang, Golin, Osberger, and Hansen, as applied to claim 40 above, and further in view of Haitsma et al. (US 2002/0178410 A1), hereinafter known as Haitsma.

61.     Stelovsky, Wang, Golin, Osberger, and Hansen teach all the features of claim 40 as demonstrated above. Stelovsky teaches means for displaying assembled visual content comprising sub-shots with music sub-clips (Column 3, Lines 27-41). Hansen teaches wherein the means for defining and selecting visual content sub-shots is such that the sub-shots are uniformly distributed within the visual content (Para. 0085-88). What Stelovsky, Wang, Golin, Osberger, and Hansen fail to teach is where the sub-shots are uniformly distributed within the visual content is further configured for selecting uniformly distributed sub-shots via evaluating normalized entropy of the sub-shots along a time line of visual content from which the sub-shots were obtained, such that displaying the assembled visual content preserves a storyline as represented by the visual content [Claim 45]. However, Haitsma teaches a hashing method for indexing video clips in a database, in which a normal distribution is calculated for video clips to determine whether they are different quality versions of the same content (Para. 0041). This is understood to be normalized entropy in the sense that the normal video quality is used to determine the similarity of sub-shots. Such a method would be used in the system and method of Stelovsky to determine whether a video clip or photograph duplicates the content of another except in quality. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have selected a uniform distribution of sub-shots along a timeline, as taught by Hansen, by analyzing the normalized entropy of the sub-shots, as taught

by Haitsma, in light of the teachings of Stelovsky, Wang, Golin, and Osberger, in order to avoid

the non-uniform selection of duplicate sub-shot content in sub-shots that have distinct data

representations due to differing quality [Claim 45].


62.     Claim 51 is are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky,

Golin, Osberger, and Hansen, in view of Borden, IV et al. (US 2003/0200105 A1), hereinafter

known as Borden IV.

63.     Stelovsky, Wang, Golin, Osberger, and Hansen teach all the features of claims 1 & 50

above. What Stelovsky, Wang, Golin, Osberger, and Hansen fail to teach is where the style

includes a day-by-day style, wherein a title is added when a new day starts before a first sub-

shot of the day to illustrate the creating of the sub-shots coming next [Claim 51]. However,

Border IV teaches a video analyzer (user's data processing device) where a user may add

metadata to photographs including the title, subject, or date of the photo (Para. 0016). The

superimposed title used by Stelovsky (10:5-6) would merely display subject or date information

of the requisite group of photographs, because this is valuable information to be conveyed in a

movie title. Therefore, it would have been obvious to one of ordinary skill in the art, at the time

the invention was made, to have one of the styles in the karaoke device of Stelovsky be a style

that displays the date in the title, as taught by Borden IV, in light of Wang, Golin, Osberger, and

Hansen, in order to make the karaoke background generated thus more informative to the user

[Claim 51].


64.     Claim 52 is are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky,

Golin, Osberger, Hansen, and Borden, IV, and further in view of Casey et al. (US 2005/0063613

A1), hereinafter known as Casey.

65.     Stelovsky, Golin, Osberger, Hansen, and Borden, IV teach all the features of claims 1,
49, & 50, as demonstrated above. What Stelovsky, Golin, Osberger, Hansen, and Borden, IV.
Fail to teach is wherein the style includes an old movie style, wherein sepia tone or grayscale
effect is added on the sub-shots [Claim 52]. However, Casey teaches where an uploaded color
photograph is grayscale or sepia tone (Para. 0036). It would be a simple matter to merely
convert the user's photographs in Stelovsky's device to grayscale, black & white, or sepia tone,
in order to filter the photograph or give it a quality of tone. Therefore, it would have been
obvious to one of ordinary skill in the art, at the time the invention was made, to have one of the
styles of Stelovsky be adding a grayscale or sepia tone to a photograph, as described in Casey,
in order to make a user's photographs more interesting [Claim 52].


***Response to Arguments***

66.     Applicant's arguments filed 7/16/2009 have been considered and are not persuasive for
the following reasons. Applicant argues at page 24 that the claim 1's "filtering" is different than
Osberger's "segmentation". However, Osberger's "segmentation" is not the temporal
segmentation of a movie described in Stelovsky; what Osberger at 3:6-15 clearly states is
splitting an image into color and luminance data for determining an attention model. The action
of splitting data and discarding the unused part is analogous to filtering the data. Further, the
process of Osberger is ultimately used for the same purpose as applicant's claim 1 recites; thus,
the argument that Osberger does not teach calculating an importance or attention index of a
shot is unpersuasive. Applicant at page 25 argues that it is not obvious in Stelovsky to
"automatically" shorten the subshots. This argument is not convincing because case precedent
has established that making a manual process like Stelovsky's "automatic" is well-known,
obvious, and of course, not patentably distinguished. Further, a number of other cited

references, such as Golin and Tsai, automatically shorten subshots. Thus applicant's argument

is not convincing. Applicant argues at page 26 that Umeda fails to preserve the story line of

visual content. However, both Umeda and Stelovsky track the dates of multimedia in metadata.

Geigel and Borden IV explicitly use this data for grouping media. It would be obvious to keep the

media in chronological order, thus preserving the story line of a media, in order to see events in

logical order (it would not make sense to see an ice cream cone eaten before it is purchased).

Thus, applicant's argument here is further unconvincing. Applicant further requests a guide to

the explicit teaching of Hansen that sub shots are selected according to a uniform distribution on

a timeline. Examiner notes that Hansen teaches at Para. 0088 that "commercials should be

presented once per minute in ten second maximum durations" in a television microchannel.

Applicant's argument is not convincing because "once per minute in ten second durations" is an

example of a uniform distribution of sub-shots of the microchannel along a timeline. Regarding

applicant's assertions that promulgated motivations in claims 17 & 18 nullify a design choice

rationale, examiner points out that an *express motivation is not given in the specification as*

*originally filed for specifically why the music is segmented according to claimed formula for*

*bounding the sub-clip's length according to: minimum length = min(max(2\*tempo,2),4) and*

*maximum length = minimum length+2 [Claim 17], or establishing the music sub-clip's length*

*within a range of 3 to 5 seconds [Claim 18].* However, Applicant has not disclosed that having

(min(max(2\*tempo,2),4) < length < min(max(2\*tempo,2),4)+2) or (3 < length < 5) seconds

specifically solves any problem or has inventive purpose; thus, it is correctly understood to be a

mere matter of design choice when compared to Stelovsky's way of dividing music subclips by,

say, 6 second increments; thus, none of applicant's arguments are persuasive.

67.     Applicant's arguments with respect to new claims 46-52 have been considered but are

moot in view of the new ground(s) of rejection. Examiner's position is that the new limitations

are mere trivialities and extra-solution activity, taught by the body of analogous literature, that

fails to make the claimed invention patentably distinguished from the cited references.


### *Conclusion*

Any inquiry concerning this communication or earlier communications from the examiner

should be directed to NIKOLAI A. GISHNOCK whose telephone number is (571)272-1420. The

examiner can normally be reached on M-F 11:00a-7:30p EST (8:00a-4:30p PST).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Xuan M. Thai can be reached on 571-272-7147. The fax phone number for the

organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent

Application Information Retrieval (PAIR) system. Status information for published applications

may be obtained from either Private PAIR or Public PAIR. Status information for unpublished

applications is available through Private PAIR only. For more information about the PAIR

system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private

PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you

would like assistance from a USPTO Customer Service Representative or access to the

automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.


9/28/2009
/N. A. G./
Examiner, Art Unit 3715

/XUAN M. THAI/
Supervisory Patent Examiner, Art Unit 3715